PATENT **MS158346.01/MSFTP184US**

CERTIFICATE OF FACSIMILE TRANSMISSION
I hereby certify that this correspondence (along with any paper referred to as being attached or enclosed) is being faxed to 571-273-8300 on the date shown below to Mail Stop Appeal Brief - Patents, Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450.

RECEIVED
CENTRAL FAX CENTER
SEP 2 0 2005

Date: 9-20-05

Himanshu S. Amin

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re patent application of:

Applicant(s):	David E. Heckerman, *et al.*	Examiner:	Wilber L. Starks
Serial No:	09/873,719	Art Unit:	2129
Filing Date:	June 4, 2001		

Title: EFFICIENT DETERMINATION OF SAMPLE SIZE TO FACILITATE BUILDING A STATISTICAL MODEL

Mail Stop Appeal Brief – Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

09/21/2005 TL0111 00000020 09873719
01 FC:1402 500.00 OP

APPEAL BRIEF

Dear Sir:

Appellants' representative submits this brief in connection with an appeal of the above-identified patent application. A credit card payment form is filed concurrently herewith in connection with all fees due regarding this appeal brief. In the event any additional fees may be due and/or are not covered by the credit card, the Commissioner is authorized to charge such fees to Deposit Account No. 50-1063 [MSFTP184US].

I. Real Party in Interest (37 C.F.R. §41.37(c)(1)(i))

The real party in interest in the present appeal is Microsoft Corporation, the assignee of the present application.

II. Related Appeals and Interferences (37 C.F.R. §41.37(c)(1)(ii))

Appellants, appellants' legal representative, and/or the assignee of the present application are not aware of any appeals or interferences which may be related to, will directly affect, or be directly affected by or have a bearing on the Board's decision in the pending appeal.

III. Status of Claims (37 C.F.R. §41.37(c)(1)(iii))

Claims 1-64 stand rejected by the Examiner. The rejection of claims 1-64 is being appealed.

IV. Status of Amendments (37 C.F.R. §41.37(c)(1)(iv))

Though claims 3-18, 21-29, 32-38, 41, 45-48, 50-51, 55-57 and 59-60 were amended after Final Office Action to correct minor informalities and to place the application in better form for appeal, the Advisory Action dated July 27, 2005 indicates that the Examiner has not entered these amendments.

V. Summary of Claimed Subject Matter (37 C.F.R. §41.37(c)(1)(v))

Independent Claim 1

Independent claim 1 recites a computer implemented system that facilitates building a statistical model for a computer readable data set, comprising, a first training algorithm that efficiently builds a rough model from a subset of the computer readable data set, an evaluation component that determines whether the subset of the computer readable data set is an appropriate subset to build a model for the computer readable data set, and a second training algorithm that builds a refined model for the computer readable data set from the subset if deemed appropriate by the evaluation component. (*See e.g.*, page 2, lines 12-21 and page 4, line 16-page 5, line 28).

Independent Claim 19

Independent claim 19 recites a computer implemented system programmed to facilitate building a statistical model, comprising, a first parameter estimation algorithm that efficiently builds a rough model from a subset of a computer readable data set based on a training policy associated therewith, and an evaluation component that determines whether the subset of data from which the rough model was built is an appropriate size for building the statistical model to characterize the data set, a second parameter estimation algorithm that builds a refined model for the data set from the subset if determined to have the appropriate size, the second parameter estimation algorithm having an associated training policy, which enables the second parameter estimation algorithm to build a more accurate model than the first parameter estimation algorithm. (*See e.g.*, page 4, line 16-page 5, line 28).

Independent Claim 30

Independent claim 30 recites a computer implemented learning curve method to facilitate building a statistical model, comprising, choosing a subset of a computer readable data set, employing a first training algorithm to build a rough model to characterize the subset, evaluating the rough model, if the rough model is unacceptable, repeatedly increasing the size of the subset of data to provide an aggregate data set, building another rough model to characterize the aggregate subset, and reevaluating the model; and if the model is acceptable, employing a second training algorithm to build a refined model based on the aggregate data set, the second training algorithm being different from the first training algorithm. (*See e.g.*, page 4, line 16-page 5, line 28).

Independent Claim 42

Independent claim 42 recites a computer-readable medium having computer-executable instructions for: choosing a subset of a computer readable data set, building a rough model to characterize the subset based on an associated training policy, evaluating the rough model, if the rough model is unacceptable, repeatedly increasing the size of the subset of data to provide an aggregate data set, building a rough model to characterize the aggregate subset based on an associated training policy, and reevaluating the rough

model, and building a refined model for the computer readable data set from the aggregate data set if the rough model is determined to be acceptable based on an associated training policy. (*See e.g.*, page 5, line 29-page 12, line 2).

Independent Claim 44

Independent claim 44 recites a computer implemented method to facilitate constructing a statistical model, comprising, separating computer readable data into holdout data and training data, determining a data subset from the training data by estimating model parameters according to a first training policy and evaluating the estimated model parameters relative to the holdout data set and repeating the estimation and evaluation of model parameters with a larger subset of the training data until an acceptable quality of the estimated model is established, and, subsequent to establishing the acceptable quality of the estimated model, using the determined data subset to improve the estimated model parameters by employing a second training policy that is more accurate than the first training policy. (*See e.g.*, page 5, line 29-page 12, line 2).

Independent Claim 53

Independent claim 53 recites computer-readable medium having computer-executable instructions for separating computer readable data into holdout data and training data, determining a data subset from the training data by estimating model parameters according to a first training policy and evaluating the estimated model parameters relative to the holdout data set and repeating the estimation and evaluation of model parameters with a next successively larger subset of the training data set until an acceptable quality of the estimated model is established, and subsequent to establishing the acceptable quality of the estimated model, using the determined data subset to improve the estimated model parameters by employing a second training policy that is more accurate than the first training policy. (*See e.g.*, page 15, line 21-page 17, line 23).

Independent Claim 54

Independent claim 54 recites a computer implemented method to facilitate constructing a statistical model, comprising: separating computer readable data into a holdout data set and a training data set, iteratively estimating model parameters for a subset of the training data set over a fixed number of iterations and evaluating the estimated model parameters relative to the holdout data set, repeating the estimation and evaluation of model parameters obtained with successively larger subsets of the training data set until an acceptable model quality is established, and after the acceptable model quality is established, iteratively estimating model parameters for the data subset, which provided the acceptable model quality, until a better quality of model is provided relative to a preceding estimation performed over the fixed number of iterations. (*See e.g.*, page 17, line 24-page 19, line 27).

Independent Claim 62

Independent claim 62 recites a computer implemented method to facilitate constructing a statistical model, comprising: separating computer readable data into a holdout data set and a training data set, iteratively estimating model parameters for a subset of the training data set until a first convergence threshold is satisfied and evaluating the estimated model parameters relative to the holdout data set, repeating the estimation and evaluation of model parameters obtained with successively larger subsets of the training data set until determining a size of data subset that provides acceptable model parameters, and after determining the size of data subset that provides acceptable model parameters, iteratively estimating model parameters for a data subset of the acceptable size until a second convergence threshold is satisfied, the second convergence threshold being less than the first convergence threshold. (*See e.g.*, page 15, line 21-page 19, line 27).

Independent Claim 63

Independent claim 63 recites a computer implemented system to facilitate building a statistical model for a computer readable data set, comprising: first means for building a rough model to characterize a subset of the computer readable data set. (*See*

e.g., page 6, lines 1-2). Independent claim 63 also recites an evaluation means for evaluating the acceptability of the rough model, the first means building another rough model for a larger subset of the data if the evaluation means determines that a prior rough model is unacceptable. (*See e.g.*, page 8, lines 13-22). Independent claim 63 further recites a second means, which is different from the first means, for building a refined model from an aggregate subset of data that yielded the rough model deemed acceptable by the evaluation means. (*See e.g.*, page 5, lines 16-28).

The means for limitations described above are identified as limitations subject to the provisions of 35 U.S.C. §112 ¶6. The structures corresponding to these limitations are identified with reference to the specification and drawings in the above-noted parentheticals.

<u>**Independent Claim 64**</u>

Independent claim 64 recites a computer implemented system to facilitate building a statistical model for a computer readable data set, comprising: first means for estimating model parameters from a subset of the computer readable data set. (*See e.g.*, page 5, line 31-page 6, line 2). Independent claim 64 also recites means for evaluating the estimated model parameters relative to a holdout set of the data set. (*See e.g.*, page 8, lines 13-22). Independent claim 64 further recites means for determining a data subset from the training data by causing the first means and the means for evaluating to respectively repeat estimation and evaluation of model parameters with a next successively larger subset of the training data set until an acceptable quality of the model parameters is established. (*See e.g.*, page 10, line 26-page 11, line 6). Additionally, independent claim 64 recites second means for estimating model parameters based on the determined data subset to provide a more accurate estimation of model parameters than the first means. (*See e.g.*, page 5, lines 16-28).

The means for limitations described above are identified as limitations subject to the provisions of 35 U.S.C. §112 ¶6. The structures corresponding to these limitations are identified with reference to the specification and drawings in the above-noted parentheticals.

VI. Grounds of Rejection to be Reviewed (37 C.F.R. §41.37(c)(1)(vi))

A. Claims 1-64 stand rejected under 35 U.S.C. §101 as it is alleged that the subject claims are directed to non-statutory subject matter.

B. Claims 1-64 stand rejected under 35 U.S.C. §112, first paragraph, because it is alleged that current case law and the MPEP require such rejection for claims that stand rejected under 35 U.S.C. §101.

C. Claims 1, 19, 30, 42 and 64 stand rejected under 35 U.S.C. §102(b) as being anticipated by Guha *et al.* (US 5,140,530).

VII. Argument (37 C.F.R. §41.37(c)(1)(vii))

A. Rejection of Claims 1-64 Under 35 U.S.C. §101

Claims 1-64 stand rejected under 35 U.S.C. §101 as it is alleged that the subject claims are directed to non-statutory subject matter. Reversal of this rejection is requested for at least the following reasons. The subject claims produce a useful, concrete and tangible result, and furthermore, the subject claims pertain to the utilization of software code to produce the useful, concrete and tangible result.

> Because the claimed process applies the Boolean principle [abstract idea] ***to produce a useful, concrete, tangible result*** ... on its face the claimed process comfortably falls within the scope of §101. *AT&T Corp. v. Excel Communications, Inc.*, 172 F.3d 1352, 1358. (Fed.Cir. 1999) (Emphasis added); *See State Street Bank & Trust Co. v. Signature Fin. Group, Inc.*, 149 F.3d 1368, 1373, 47 USPQ2d 1596, 1601 (Fed.Cir.1998). The inquiry into patentability requires an examination of the contested claims to see if the claimed subject matter, as a whole, is a disembodied mathematical concept representing nothing more than a "law of nature" or an "abstract idea," or if the mathematical concept has been ***reduced to some practical application rendering it "useful."*** *AT&T* at 1357 *citing In re Alappat*, 33 F.3d 1526, 31 1544, 31 U.S.P.Q.2D (BNA) 1545, 1557 (Fed. Cir. 1994) (Emphasis added) (holding that more than an abstract idea was claimed because the claimed invention as a whole was directed toward forming

a specific machine that produced the useful, concrete, and tangible result of a smooth waveform display).

The subject invention, as evinced by independent claims 1, 19, 30, 42, 44, 53, 54, 62, 63, and 64, produces a useful, concrete, and tangible result. Independent claim 1 (and similarly independent claims 19, 30, 42, 44, 53, 54, 62, 63, and 64) recites a first training algorithm that efficiently builds a rough model from a subset of the computer readable data set; an evaluation component that determines whether the subset of the computer readable data set is an appropriate subset to build a model for the computer readable data set; and a second training algorithm that builds a refined model for the computer readable data set from the subset if deemed appropriate by the evaluation component.

The appellants' claimed invention yields a number of useful, concrete, and tangible results. In particular, the subject claims recite that a refined model for the computer readable data set is built based on an appropriate subset from the computer readable data set. This refined model is a useful, concrete and tangible result. For example, one would appreciate that the refined model can be employed in connection with clustering, data mining, *etc.* Additionally, the appellants' claims recite that an appropriate subset from which to build a model is determined. The determination of the appropriate subset is a useful, tangible, and concrete result since it enables identifying a subset from which to build the refined model that provides for a balance between accuracy and efficiency associated with model generation.

In the Office Action dated May 18, 2005 the Examiner asserted "that Applicant manipulated a set of abstract 'computer readable data sets' to solve purely algorithmic problems in the abstract." (*See* page. 6). Appellants' representative disagrees with such contention. Similar to the result produced in *State Street Bank & Trust Co. v. Signature Fin. Group, Inc.*, 149 F.3d 1368, the manipulation of computer readable data sets (*e.g.*, building a rough model from a subset, evaluating the subset to determine whether it is appropriate, building a refined model based on the appropriate subset, ...) constitutes a practical application because it produces useful, concrete and tangible results – namely, a refined model of the computer readable data set and a determination of an appropriate subset from which to build the refined model. Thus, the subject claims are not directed to

manipulating an abstract idea since the claims relate to a practical application that is useful, concrete and tangible.

Moreover, the Court of Appeals for the Federal Circuit stated in *Eolas Techs., Inc. v. Microsoft Corp.*, 399 F.3d 1325 (Fed. Cir. 2005):

> Title 35, section 101, explains that an invention includes "any new and useful process, machine, manufacture or composition of matter." ... Without question, ***software code alone qualifies as an invention eligible for patenting under these categories***, at least as processes. *Id.* at 1338 (emphasis added).

The subject claims clearly pertain to software code comprising a first training algorithm that efficiently builds a rough model from a subset of the computer readable data set; an evaluation component that determines whether the subset of the computer readable data set is an appropriate subset to build a model for the computer readable data set; and a second training algorithm that builds a refined model for the computer readable data set from the subset if deemed appropriate by the evaluation component. The fact that (i) the subject claims elicit a useful, concrete and tangible result, and (ii) the result so elicited is the produced *via* execution of software code, leads one to conclude that the Examiner's rejection under 35 U.S.C. §101 is clearly erroneous.

In view of at least the foregoing, it is readily apparent that the claimed invention reduces to a practical application that produces a useful, concrete, tangible result, pursuant to *AT&T Corp. v. Excel Communications, Inc.*, 172 F.3d 1352, 1358 (Fed. Cir. 1999), and further that the result so produced is provided by the execution of software code, which according to *Eolas Techs., Inc. v. Microsoft Corp.*, 399 F.3d 1325 (Fed. Cir. 2005) is patentable *per se*. Thus, contrary to the Examiner's assertions, it is believed that the subject claims are directed to statutory subject matter pursuant to 35 U.S.C. §101. Accordingly, this rejection should be reversed.

B. <u>Rejection of Claims 1-64 Under 35 U.S.C. §112</u>

Claims 1-64 stand rejected under 35 U.S.C. §112, first paragraph, because it is alleged that current case law and the MPEP require such rejection for claims that stand

rejected under 35 U.S.C. §101. It is believed that this rejection is improper and should be reversed for at least the following reasons. The rejection of claims 1-64 under 35 U.S.C. §101 should be reversed pursuant to the aforementioned comments rendering the subject rejection moot. Accordingly, reversal of this rejection is requested.

C. Rejection of Claims 1, 19, 30, 42 and 64 Under 35 U.S.C. §102(b)

Claims 1, 19, 30, 42 and 64 stand rejected under 35 U.S.C. §102(b) as being anticipated by Guha *et al.* (US 5,140,530). Reversal of this rejection is requested for at least the following reasons. Guha *et al.* does not disclose or suggest all aspects of the subject claims.

> A single prior art reference anticipates a patent claim only if it ***expressly or inherently describes each and every limitation set forth in the patent claim***. *Trintec Industries, Inc. v. Top-U.S.A. Corp.*, 295 F.3d 1292, 63 USPQ2d 1597 (Fed. Cir. 2002); *See Verdegaal Bros. v. Union Oil Co. of California,* 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987). The ***identical invention must be shown in as complete detail as is contained in the ... claim***. *Richardson v. Suzuki Motor Co.*, 868 F.2d 1226, 9 USPQ2d 1913, 1920 (Fed. Cir. 1989) (emphasis added).

The subject claims relate to systems and methods that facilitate building a model to characterize data based on an appropriately sized subset of the computer readable data set. In particular, independent claim 1 (and similarly independent claims 19, 30, 42, and 64) recites an evaluation component that determines whether the subset of the computer readable data set is an appropriate subset to build a model for the computer readable data set and a second training algorithm that builds a refined model for the computer readable data set from the subset if deemed appropriate. Guha *et al.* fails to disclose or suggest such claimed aspects.

More particularly, Guha *et al.* does not disclose or suggest employing a subset of the computer readable data set as recited in the subject claims. The Final Office Action asserts that "the 'network blueprints' shown in Fig. 2 are the design parameters (or the 'subsets' of 'computer readable data'...) being used to build the candidate models in the

genetically evolving population. (*See* Final Office Action dated May 18, 2005, page 11). Appellants' representative avers to the contrary. The blueprints as disclose in Guha *et al.* are bit stream designs for different neural networks. (*See* col. 2, lines 63-66). The blueprints can specify genetic algorithm parameters that determine how the genetic operators are used to construct network structures and an evaluation function that determines the fitness of a network for a specific application. (*See* col. 3, lines 55-61). However, Guha *et al.* is silent regarding the blueprint being a subset from a data set which is to be modeled. The appellants' claims instead relate to employing a subset from a data set to build a model that represents the data set; hence, a portion of or an entire data set is employed in connection with the modeling the data set. Thus, Guha *et al.* fails to anticipate or suggest such claimed aspects.

Furthermore, Guha *et al.* does not anticipate or suggest an evaluation component that determines whether the subset of the computer readable data set is an appropriate subset to build a model for the computer readable data set as claimed. The Final Office Action contends that "the box that performs network performance evaluation in Fig. 2" discloses such aspects since "the genetic algorithm uses this process to determine whether the specific network blueprints ... are appropriate subsets to build a model for the computer readable data set." (*See* Final Office Action dated May 18, 2005, page 12). Appellants' representative respectfully disagrees with such contentions. Guha *et al.* discloses that the fitness of a network can be determined by the evaluation function. (*See* col. 3, lines 59-61). However, Guha *et al.* does not evaluate whether a subset from a data set which was utilized to build a model is an appropriate subset since the blueprints are not subsets of the data sets as noted previously. Thus, Guha *et al.* fails to teach or suggest appellants' invention as claimed.

Moreover, Guha *et al.* does not teach or suggest a second training algorithm that builds a refined model for the computer readable data set from the subset if deemed appropriate as recited in the subject claims. The Final Office Action contends that the "'second training algorithm' ... is the algorithm that is used to take the untrained network, at the bottom of Fig. 2, into a trained state, at the bottom-right of Fig. 2." (*See* Final Office Action dated May 18, 2005, page 13). Appellants' representative disagrees with such contentions. Guha *et al.* updates blueprints in a cyclical manner as depicted in

Fig. 2. Fig. 2 illustrates that an untrained network is trained, and then the trained network is evaluated to determine the blueprint fitness. Thus, Guha *et al.* fails to anticipate or suggest that a second training algorithm builds a refined model ***from the subset if deemed appropriate.***

In addition, it is emphasized that the standard by which anticipation is to be measured is ***strict identity*** between the cited document and the invention as claimed, not mere equivalence or similarity. *See, Richardson* at 9 USPQ2d 1913, 1920. This means that in order to establish anticipation under 35 U.S.C. §102, the single document cited must not only expressly or inherently describe each and every limitation set forth in the patent claim, but also the identical invention must be shown in as complete detail as is contained in the claim. The fact that Guha *et al.* (a) does not employ a subset of the computer readable data set, but rather discloses the utilization of blueprints <u>without</u> actually disclosing or informing one of ordinary skill in the art that the blue prints so disclosed constitutes a subset of the data set to be modeled; (b) does not provide an evaluation component that determines whether the subset of the computer readable data set is an appropriate subset to build a model for the computer readable data set; and (c) does not disclose or suggest a second training algorithm that builds a refined model for the computer readable data set from the subset if deemed appropriate, indicates that the cited document does not provide an invention identical to that recited in the subject claims.

Further, it is believed that the Examiner has failed to fully satisfy his burden under MPEP §§707.07(i) and 2106, which state that in "every Office action, each pending claim should be mentioned by number, and its treatment or status given", (*See* MPEP §707.07(i), and even though claims may be perceived to fall within the ambit of 35 U.S.C. §§ 101 and 112, first paragraph in their entirety, that this "should not preclude complete examination of the application for satisfaction of all other conditions of patentability." (*See* MPEP 2106). It is submitted that in both the Office Action dated December 2, 2004, and the Final Office Action dated May 18, 2005, the Examiner, while rejecting the subject claims in their entirety under 35 U.S.C. §§ 101 and 112, first paragraph, has nevertheless not satisfied the obligation imposed by the aforementioned sections of the MPEP under 35 U.S.C. §§ 102 and 103.

In view of at least the foregoing, it is apparent that Guha *et al.* does not disclose or suggest the subject invention as recited in claims 1, 19, 30, 42, and 64. Further, in light of the Examiner's failure to specifically address and give indication of the status of claims 2-18, 20-29, 31-41, and 43, which respectively depend from independent claims 1, 19, 30, and 42, as well as claims 44-63, it is therefore believed that these claims are in condition for allowance. Accordingly, this rejection should be reversed.

D. Conclusion

For at least the above reasons, the claims currently under consideration are believed to be patentable over the cited references. Accordingly, it is respectfully requested that the rejections of claims 1-64 be reversed.

If any additional fees are due in connection with this document, the Commissioner is authorized to charge those fees to Deposit Account No. 50-1063.

Respectfully submitted,
AMIN & TUROCY, LLP

Himanshu S. Amin
Reg. No. 40,894

AMIN & TUROCY, LLP
24th Floor, National City Center
1900 East 9th Street
Telephone: (216) 696-8730
Facsimile: (216) 696-8731

VIII. Claims Appendix (37 C.F.R. §41.37(c)(1)(viii))

1. A computer implemented system that facilitates building a statistical model for a computer readable data set, comprising:

a first training algorithm that efficiently builds a rough model from a subset of the computer readable data set;

an evaluation component that determines whether the subset of the computer readable data set is an appropriate subset to build a model for the computer readable data set; and

a second training algorithm that builds a refined model for the computer readable data set from the subset if deemed appropriate by the evaluation component.

2. The system of claim 1, further comprising a data scheduler which, based on a data policy, controls the size of subsets for which the first training algorithm is applied.

3. The system of claim 2, wherein the data scheduler increases the size of the subset to provide a larger aggregate subset of the data set if the rough model is unacceptable, the first training algorithm efficiently builds the rough model for each larger aggregate subset of the data until the evaluation component determines the resulting rough model to be acceptable.

4. The system of claim 3, wherein the acceptability of each rough model is determined based on a stopping criterion functionally related to an expected incremental benefit and a cost associated with increasing the size of the aggregate subset of the data set.

5. The system of claim 4, wherein the cost of the stopping criterion is functionally related to at least one of time associated with evaluating an aggregate data subset of increased size and size of the aggregated subset of the data.

6. The system of claim 4, wherein the stopping criterion is defined by

$$\left(\frac{l(D_{HO} \mid \theta(D_n)) - l(D_{HO} \mid \theta(D_{n-1}))}{l(D_{HO} \mid \theta(D_n)) - l(D_{HO} \mid \theta_{BASE}(D_n))}\right) \frac{1}{c_1(I_1 - \bar{J}_n) \mid \Delta D_{n+1} \mid + c_2(I_1 - \bar{J}_n) + c_1\bar{J}_n \mid D_{n+1} \mid + c_2\bar{J}_n + c_3} < \lambda$$

where

$l(D_{HO}|\theta(D_n))$ is a log likelihood for holdout data evaluated for the model built by the first training algorithm on a current subset of the training data set,

$l(D_{HO}|\theta(D_{n-1}))$ is a log likelihood for holdout data evaluated for the model built by the first training algorithm on a previous subset of the training data set,

$l(D_{HO}|\theta_{base}(D_n))$ is a log likelihood for holdout data evaluated for a base model,

c_1, c_2, and c_3 are constants determined based on application of the second training algorithm relative to a first subset of the data set,

I_1 is a number of iterations for the second training algorithm, when applied to the first subset,

$\bar{J}_n = \frac{1}{n}\sum_{i=1}^{n} J_i$, and

J_i is the number of iterations for the first training algorithm when applied to a data subset D_i,

$|D_{n+1}|$ is the size of data set D_{n+1},

$|\Delta D_{n+1}|$ is the increment in size $|D_{n+1}| - |D_n|$,

λ is a user determined stopping threshold .

7. The system of claim 4, wherein the stopping criterion is defined by

$$\left(\frac{l(D_{HO}|\theta(D_n))-l(D_{HO}|\theta(D_{n-1}))}{l(D_{HO}|\theta(D_n))+\delta-l(D_{HO}|\theta_{BASE}(D_n))}\right)\frac{1}{c_1(I_1-\bar{J}_n)|\Delta D_{n+1}|+c_2(I_1-\bar{J}_n)+c_1\bar{J}_n|D_{n+1}|+c_2\bar{J}_n+c_3}<\lambda$$

where

$l(D_{HO}|\theta(D_n))$ is a log likelihood for holdout data evaluated for the model built by the first training algorithm on a current subset of the training data set,

$l(D_{HO}|\theta(D_{n-1}))$ is a log likelihood for holdout data evaluated for the model built by the first training algorithm on a previous subset of the training data set,

$l(D_{HO}|\theta_{base}(D_n))$ is a log likelihood for holdout data evaluated for a base model,

δ is an offset associated with a difference in log likelihood for holdout data when evaluated for models built on a first subset of the training data set by the respective first and second training algorithms,

c_1, c_2, and c_3 are constants determined based on application of the second training algorithm relative to a first subset of the data set,

I_1 is a number of iterations for the second training algorithm, when applied to the first subset,

$\bar{J}_n = \frac{1}{n}\sum_{i=1}^{n} J_i$, and

J_i is the number of iterations for the first training algorithm when applied to a data subset D_i,

$|D_{n+1}|$ is the size of data set D_{n+1},

$|\Delta D_{n+1}|$ is the increment in size $|D_{n+1}| - |D_n|$, and

λ is a user determined stopping threshold.

8. The system of claim 1, wherein the first training algorithm further comprises an iterative algorithm, which builds the rough model for the subset of the data set according to an associated training policy.

9. The system of claim 8, wherein the first training algorithm further comprises an associated training policy that defines parameter initialization of the first training algorithm for each subset of the data set.

10. The system of claim 9, wherein the training policy associated with the first training algorithm further controls parameter initialization of the first training algorithm, such that at least some of the parameters computed for a previous subset of the data are employed to initialize the first training algorithm for a subsequent larger aggregate subset of the data.

11. The system of claim 9, wherein the first training algorithm is initialized by the same parameter values for each subset of the data subset.

12. The system of claim 9, wherein the training policy sets the iterative algorithm to perform a fixed number of at least one iteration.

13. The system of claim 12, wherein the training policy sets the iterative algorithm to perform a single iteration.

14. The system of claim 12, wherein the second training algorithm further comprises an iterative algorithm that operates according to an associated training policy, so as to produce a more accurate model for the appropriate subset of the data set than the first training algorithm.

15. The system of claim 14, wherein the iterative algorithm associated with at least one of the first and second training algorithms is an Expectation and Maximization algorithm.

16. The system of claim 8, wherein the training policy associated with the iterative algorithm of the first training algorithm controls the iterative algorithm to run until an associated convergence criterion is satisfied.

17. The system of claim 16, wherein second training algorithm further comprises an iterative algorithm, which builds the refined model for the appropriate subset of the data set according to an associated training policy.

18. The system of claim 17, wherein the training policy associated with the iterative algorithm of the second training algorithm controls the respective iterative algorithm to run until an associated convergence criterion is satisfied, wherein the convergence criterion associated with the second training algorithm provides improved model quality relative to the convergence criterion associated with the first training algorithm.

19. A computer implemented system programmed to facilitate building a statistical model, comprising:

a first parameter estimation algorithm that efficiently builds a rough model from a subset of a computer readable data set based on a training policy associated therewith; and

an evaluation component that determines whether the subset of data from which the rough model was built is an appropriate size for building the statistical model to characterize the data set;

a second parameter estimation algorithm that builds a refined model for the data set from the subset if determined to have the appropriate size, the second parameter estimation algorithm having an associated training policy, which enables the second parameter estimation algorithm to build a more accurate model than the first parameter estimation algorithm.

20. The system of claim 19, further comprising a data scheduler that increases the size of the subset of the data set to provide a larger aggregate subset of the data set if the rough model is unacceptable, the first parameter estimation algorithm efficiently builds a rough model for each larger aggregate subset until a resulting rough model built therefrom is determined to be acceptable.

21. The system of claim 19, wherein the first parameter estimation algorithm further comprises an iterative algorithm that builds the rough model for each subset of the data set according to the associated training policy.

22. The system of claim 21, wherein the training policy for the first parameter estimation algorithm is operative to control parameter initialization for the first parameter estimation algorithm, such that at least some of the parameters computed for a previous subset of the data are employed to initialize the first parameter estimation algorithm for a subsequent larger aggregate subset of the data set.

23. The system of claim 21, wherein the first parameter estimation algorithm is initialized by the same parameter values for each subset of the data subset.

24. The system of claim 21, wherein the training policy associated with first parameter estimation algorithm controls the iterative algorithm of the first parameter estimation algorithm to perform a fixed number of at least one iteration, the second training algorithm further comprising an iterative algorithm, which is operative to perform a greater number of iterations than the iterative algorithm of the first training algorithm based on a training policy associated with the second parameter estimation algorithm.

25. The system of claim 21, wherein the training policy associated with the iterative algorithm of the first parameter estimation algorithm controls the iterative algorithm to run until an associated convergence threshold is satisfied, wherein the second training algorithm further comprises an iterative algorithm, the training policy associated with the iterative algorithm of the second parameter estimation algorithm being operative to control the respective iterative algorithm to run until an associated convergence threshold is satisfied, the convergence threshold associated with the second parameter estimation algorithm is less than the convergence threshold associated with the first parameter estimation algorithm.

26. The system of claim 19, wherein the evaluation component determines whether the subset of data for which the rough model was built is an appropriate size based on a stopping criterion, which is functionally related to an expected incremental benefit and an expected incremental cost associated with increasing size of the subset of data.

27. The system of claim 26, wherein the cost of the stopping criterion is functionally related to at least one of time associated with evaluating the model for a larger subset of data and size of the larger subset of the data.

28. The system of claim 26, wherein the stopping criterion is defined by

$$\left(\frac{l(D_{HO}\mid\theta(D_n))-l(D_{HO}\mid\theta(D_{n-1}))}{l(D_{HO}\mid\theta(D_n))-l(D_{HO}\mid\theta_{BASE}(D_n))}\right)\frac{1}{c_1(I_1-\bar{J}_n)\mid\Delta D_{n+1}\mid+c_2(I_1-\bar{J}_n)+c_1\bar{J}_n\mid D_{n+1}\mid+c_2\bar{J}_n+c_3}<\lambda$$

where

$l(D_{HO}|\theta(D_n))$ is a log likelihood for holdout data evaluated for the model built by the first training algorithm on a current subset of the training data set,

$l(D_{HO}|\theta(D_{n-1}))$ is a log likelihood for holdout data evaluated for the model built by the first training algorithm on a previous subset of the training data set,

$l(D_{HO}|\theta_{base}(D_n))$ is a log likelihood for holdout data evaluated for a base model,

c_1, c_2, and c_3 are constants determined based on application of the second parameter estimation algorithm relative to a first subset of the data set,

I_1 is a number of iterations for the second parameter estimation algorithm, when applied to the first subset,

$\bar{J}_n=\frac{1}{n}\sum_{i=1}^{n}J_i$, and

J_i is the number of iterations for the first parameter estimation algorithm when applied to a data subset D_i,

$|D_{n+1}|$ is the size of data set D_{n+1},

$|\Delta D_{n+1}|$ is the increment in size $|D_{n+1}| - |D_n|$, and

λ is a user determined stopping threshold.

29. The system of claim 26, wherein the stopping criterion is defined by

$$\left(\frac{l(D_{HO}\,|\,\theta(D_n)) - l(D_{HO}\,|\,\theta(D_{n-1}))}{l(D_{HO}\,|\,\theta(D_n)) + \delta - l(D_{HO}\,|\,\theta_{BASE}(D_n))}\right)\frac{1}{c_1(I_1-\bar{J}_n)\,|\,\Delta D_{n+1}\,| + c_2(I_1-\bar{J}_n) + c_1\bar{J}_n\,|\,D_{n+1}\,| + c_2\bar{J}_n + c_3} < \lambda$$

where

$l(D_{HO}|\theta(D_n))$ is a log likelihood for holdout data evaluated for the model built by the first training algorithm on a current subset of the training data set,

$l(D_{HO}|\theta(D_{n-1}))$ is a log likelihood for holdout data evaluated for the model built by the first training algorithm on a previous subset of the training data set,

$l(D_{HO}|\theta_{base}(D_n))$ is a log likelihood for holdout data evaluated for a base model,

δ is an offset associated with a difference in log likelihood for holdout data when evaluated for models built on a first subset of the training data set by the respective first and second training algorithms,

c_1, c_2, and c_3 are constants determined based on application of the second parameter estimation algorithm relative to a first data subset of the data set,

I_1 is a number of iterations for the second parameter estimation algorithm, when applied to a first data subset,

$\bar{J}_n = \frac{1}{n}\sum_{i=1}^{n} J_i$, and

J_i is the number of iterations for the first parameter estimation algorithm when applied to a data subset D_i,

$|D_{n+1}|$ is the size of data set D_{n+1},

$|\Delta D_{n+1}|$ is the increment in size $|D_{n+1}| - |D_n|$, and

λ is a user determined stopping threshold.

30. A computer implemented learning curve method to facilitate building a statistical model, comprising:

choosing a subset of a computer readable data set;

employing a first training algorithm to build a rough model to characterize the subset;

evaluating the rough model;

if the rough model is unacceptable, repeatedly increasing the size of the subset of data to provide an aggregate data set, building another rough model to characterize the aggregate subset, and reevaluating the model; and

if the model is acceptable, employing a second training algorithm to build a refined model based on the aggregate data set, the second training algorithm being different from the first training algorithm.

31. The method of claim 30, further comprising determining the acceptability of each rough model based on a stopping criterion functionally related to an expected incremental benefit and an expected incremental cost associated with increasing the size of the aggregate subset of the data set.

32. The system of claim 31, wherein the cost of the stopping criterion is functionally related to at least one of time associated with evaluating an aggregate data subset of increased size and size of the aggregate subset of the data.

33. The system of claim 31, wherein the stopping criterion is defined by

$$\left(\frac{l(D_{HO} \mid \theta(D_n)) - l(D_{HO} \mid \theta(D_{n-1}))}{l(D_{HO} \mid \theta(D_n)) - l(D_{HO} \mid \theta_{BASE}(D_n))}\right) \frac{1}{c_1(I_1 - \bar{J}_n) \mid \Delta D_{n+1} \mid + c_2(I_1 - \bar{J}_n) + c_1 \bar{J}_n \mid D_{n+1} \mid + c_2 \bar{J}_n + c_3} < \lambda$$

where

$l(D_{HO}|\theta(D_n))$ is a log likelihood for holdout data evaluated for the model built by the first training algorithm on a current subset of the training data set,

$l(D_{HO}|\theta(D_{n-1}))$ is a log likelihood for holdout data evaluated for the model built by the first training algorithm on a previous subset of the training data set,

$l(D_{HO}|\theta_{base}(D_n))$ is a log likelihood for holdout data evaluated for a base model,

c_1, c_2, and c_3 are constants determined based on application of the second parameter estimation algorithm relative to a first subset of the data set,

I_1 is a number of iterations for the second parameter estimation algorithm, when applied to the first subset,

$$\bar{J}_n = \frac{1}{n}\sum_{i=1}^{n} J_i \text{, and}$$

J_i is a number of iterations for the first parameter estimation algorithm when applied to a data subset D_i,

$|D_{n+1}|$ is a size of data set D_{n+1},

$|\Delta D_{n+1}|$ is an increment in size $|D_{n+1}| - |D_n|$, and

λ is a user determined stopping threshold.

34. The system of claim 31, wherein the stopping criterion is defined by

$$\left(\frac{l(D_{HO}|\theta(D_n)) - l(D_{HO}|\theta(D_{n-1}))}{l(D_{HO}|\theta(D_n)) + \delta - l(D_{HO}|\theta_{BASE}(D_n))}\right)\frac{1}{c_1(I_1-\bar{J}_n)|\Delta D_{n+1}| + c_2(I_1-\bar{J}_n) + c_1\bar{J}_n|D_{n+1}| + c_2\bar{J}_n + c_3} < \lambda$$

where

$l(D_{HO}|\theta(D_n))$ is a log likelihood for holdout data evaluated for the model built by the first training algorithm on a current subset of the training data set,

$l(D_{HO}|\theta(D_{n-1}))$ is a log likelihood for holdout data evaluated for the model built by the first training algorithm on a previous subset of the training data set,

$l(D_{HO}|\theta_{base}(D_n))$ is a log likelihood for holdout data evaluated for a base model,

δ is an offset associated with the difference in log likelihood for holdout data when evaluated for models built on a first subset of the training data set by the respective first and second training algorithms,

c_1, c_2, and c_3 are constants determined based on application of the second parameter estimation algorithm relative to a first data subset of the data set,

I_1 is a number of iterations for the second parameter estimation algorithm, when applied to a first data subset,

$\bar{J}_n = \frac{1}{n}\sum_{i=1}^{n} J_i$, and

J_i is a number of iterations for the first parameter estimation algorithm when applied to a data subset D_i,

$|D_{n+1}|$ is a size of data set D_{n+1},

$|\Delta D_{n+1}|$ is an increment in size $|D_{n+1}| - |D_n|$, and

λ is a user determined stopping threshold.

35. The method of claim 30, wherein the first training algorithm is more computationally efficient than the second training algorithm.

36. The method of claim 30, wherein each instance of model building repeated until obtaining an acceptable model by the first training algorithm employs more efficient and less accurate model building than model building employed by the second training algorithm that occurs after obtaining the acceptable model.

37. The method of claim 36, wherein each instance of model building repeated until obtaining an acceptable model employs the first training algorithm as an iterative algorithm that is run to a first convergence criterion, the second training algorithm employing an iterative algorithm that is run to a second convergence criterion, which demands more iterations than the first convergence criterion in order to obtain convergence, so that the refined model is more accurate than the rough model built by the first training algorithm.

38. The method of claim 36, wherein each instance of model building repeated until obtaining an acceptable model employs an iterative algorithm having a fixed number of at least one iteration, the second training algorithm employing an iterative algorithm having a greater number of iterations than the fixed number.

39. The method of claim 30, further comprising controlling parameter initialization employed in each instance of building a model for the aggregate data set prior to obtaining an acceptable model.

40. The method of claim 39, further comprising initializing the first training algorithm by the same parameter values for each subset.

41. The method of claim 39, wherein the controlling further comprises reusing at least some of the parameters computed from a previous instance of model building to initialize a subsequent instance of model building for a subsequent larger aggregate data set prior to obtaining an acceptable model.

42. A computer-readable medium having computer-executable instructions for:

choosing a subset of a computer readable data set;

building a rough model to characterize the subset based on an associated training policy;

evaluating the rough model;

if the rough model is unacceptable, repeatedly increasing the size of the subset of data to provide an aggregate data set, building a rough model to characterize the aggregate subset based on an associated training policy, and reevaluating the rough model; and

building a refined model for the computer readable data set from the aggregate data set if the rough model is determined to be acceptable based on an associated training policy.

43. The method of claim 42, further comprising determining the acceptability of the model based on an expected incremental benefit relative to an expected incremental cost associated with increasing the size of the aggregate data set.

44. A computer implemented method to facilitate constructing a statistical model, comprising:

separating computer readable data into holdout data and training data;

determining a data subset from the training data by estimating model parameters according to a first training policy and evaluating the estimated model parameters relative to the holdout data set and repeating the estimation and evaluation of model parameters with a larger subset of the training data until an acceptable quality of the estimated model is established; and,

subsequent to establishing the acceptable quality of the estimated model, using the determined data subset to improve the estimated model parameters by employing a second training policy that is more accurate than the first training policy.

45. The method of claim 44, wherein each estimation of model parameters repeated until the acceptable quality of the estimated model is established further comprises employing an iterative algorithm that is run until a first convergence criterion is satisfied, the estimation of model parameters using the determined data subset further comprising an iterative algorithm that is run until a second convergence criterion is satisfied, which is operative to provide a better quality of model than the first convergence criterion.

46. The system of claim 45, wherein the first convergence criterion causes the associated iterative algorithm to run until a first convergence threshold is satisfied, wherein the second convergence criterion causes the associated iterative algorithm to run until a second convergence threshold is satisfied, the second convergence threshold being less than the first convergence threshold.

47. The method of claim 45, wherein at least one of the iterative algorithm run to the first convergence criterion and the iterative algorithm run to the second convergence criterion is an Expectation and Maximization algorithm.

48. The method of claim 44, wherein each estimation of model parameters repeated until the acceptable quality of the estimated model is established employs an iterative algorithm having a fixed number of at least one iteration, the estimation of model parameters using the determined data subset further employing an iterative algorithm having a greater number of iterations than the fixed number.

49. The method of claim 44, further comprising controlling parameter initialization employed in each estimation of model parameters repeated until determining an acceptable size for the determined data subset.

50. The method of claim 44, wherein the controlling further comprises reusing at least some of the parameters computed from a previous estimation of model parameters to initialize a subsequent estimation of model parameters for a next larger subset of the training set.

51. The method of claim 44, wherein each estimation of model parameters repeated until the acceptable quality of the estimated model is established further comprises initializing the first training algorithm by the same parameter values.

52. The method of claim 44, further comprising determining the acceptability of the estimated model based on an expected incremental benefit relative to a cost associated with increasing the size of the subset of the data set.

53. A computer-readable medium having computer-executable instructions for:

separating computer readable data into holdout data and training data;

determining a data subset from the training data by estimating model parameters according to a first training policy and evaluating the estimated model parameters relative to the holdout data set and repeating the estimation and evaluation of model parameters with a next successively larger subset of the training data set until an acceptable quality of the estimated model is established; and

subsequent to establishing the acceptable quality of the estimated model, using the determined data subset to improve the estimated model parameters by employing a second training policy that is more accurate than the first training policy.

54. A computer implemented method to facilitate constructing a statistical model, comprising:

separating computer readable data into a holdout data set and a training data set;

iteratively estimating model parameters for a subset of the training data set over a fixed number of iterations and evaluating the estimated model parameters relative to the holdout data set;

repeating the estimation and evaluation of model parameters obtained with successively larger subsets of the training data set until an acceptable model quality is established; and

after the acceptable model quality is established, iteratively estimating model parameters for the data subset, which provided the acceptable model quality, until a better quality of model is provided relative to a preceding estimation performed over the fixed number of iterations.

55. The method of claim 54, wherein at least one of the iterative estimations employs an Expectation and Maximization algorithm.

56. The method of claim 54, wherein the estimation that occurs after the acceptable model quality is established, further comprises employing an iterative algorithm having a greater number of iterations than the fixed number.

57. The method of claim 54, wherein the estimation of model parameters after the acceptable model quality has been established further comprises employing an iterative algorithm that is run until a convergence criterion is satisfied, which is operative to provide a better quality of model with the data subset than a preceding estimation employing the fixed number of iterations.

58. The method of claim 54, further comprising controlling parameter initialization for each estimation of model parameters that occurs before the acceptable model quality has been established.

59. The method of claim 58, wherein each iterative estimation until the acceptable model quality is established further comprises initializing the first training algorithm by the same parameter values.

60. The method of claim 58, wherein the controlling further comprises reusing at least some of the parameters obtained in a previous estimation of model parameters to initialize a subsequent estimation of model parameters for a next larger subset of the training data set.

61. The method of claim 54, further comprising determining the acceptability of the model based on an expected incremental benefit relative to an expected incremental cost associated with an increase in size of each larger training subset of the data set.

62. A computer implemented method to facilitate constructing a statistical model, comprising:

separating computer readable data into a holdout data set and a training data set;

iteratively estimating model parameters for a subset of the training data set until a first convergence threshold is satisfied and evaluating the estimated model parameters relative to the holdout data set;

repeating the estimation and evaluation of model parameters obtained with successively larger subsets of the training data set until determining a size of data subset that provides acceptable model parameters; and

after determining the size of data subset that provides acceptable model parameters, iteratively estimating model parameters for a data subset of the acceptable

size until a second convergence threshold is satisfied, the second convergence threshold being less than the first convergence threshold.

63. A computer implemented system to facilitate building a statistical model for a computer readable data set, comprising:

first means for building a rough model to characterize a subset of the computer readable data set;

evaluation means for evaluating the acceptability of the rough model, the first means building another rough model for a larger subset of the data if the evaluation means determines that a prior rough model is unacceptable; and

second means, which is different from the first means, for building a refined model from an aggregate subset of data that yielded the rough model deemed acceptable by the evaluation means.

64. A computer implemented system to facilitate building a statistical model for a computer readable data set, comprising:

first means for estimating model parameters from a subset of the computer readable data set;

means for evaluating the estimated model parameters relative to a holdout set of the data set;

means for determining a data subset from the training data by causing the first means and the means for evaluating to respectively repeat estimation and evaluation of model parameters with a next successively larger subset of the training data set until an acceptable quality of the model parameters is established; and

second means for estimating model parameters based on the determined data subset to provide a more accurate estimation of model parameters than the first means.

IX. Evidence Appendix (37 C.F.R. §41.37(c)(1)(ix))

None.

X. Related Proceedings Appendix (37 C.F.R. §41.37(c)(1)(x))

None.